Supplementary information for

Two DNA-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise

Maya Dadiani[1], David van Dijk[1], Barak Segal[1], Yair Field[1], Gil Ben-Artzi[1], Tali Raveh-Sadka[1], Michal Levo[1], Irene Kaplow[1], Adina Weinberger[1] and Eran Segal[1,*]

[1] Department of Computer Science and Applied Mathematics, and Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

[*] Correspondence should be addressed to: eran.segal@weizmann.ac.il

Contents

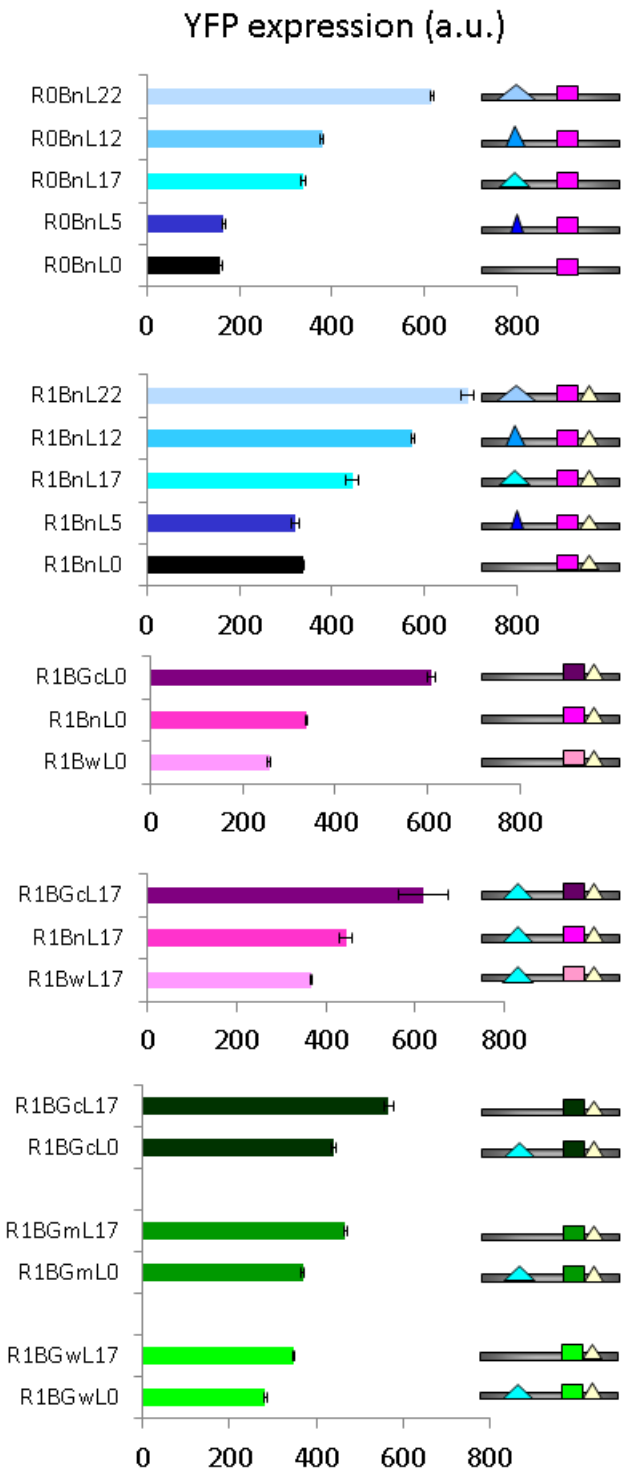

Figure S1. Promoter variants and expression levels. Shown are the promoter variants used in this study and the median YFP expression of each variant, as measured by flow cytometry. Note the increase in expression with both the lengthening of poly(dA:dT) tracts (upper two promoter sets) and the strengthening of transcription factor binding sites (four bottom promoter sets).

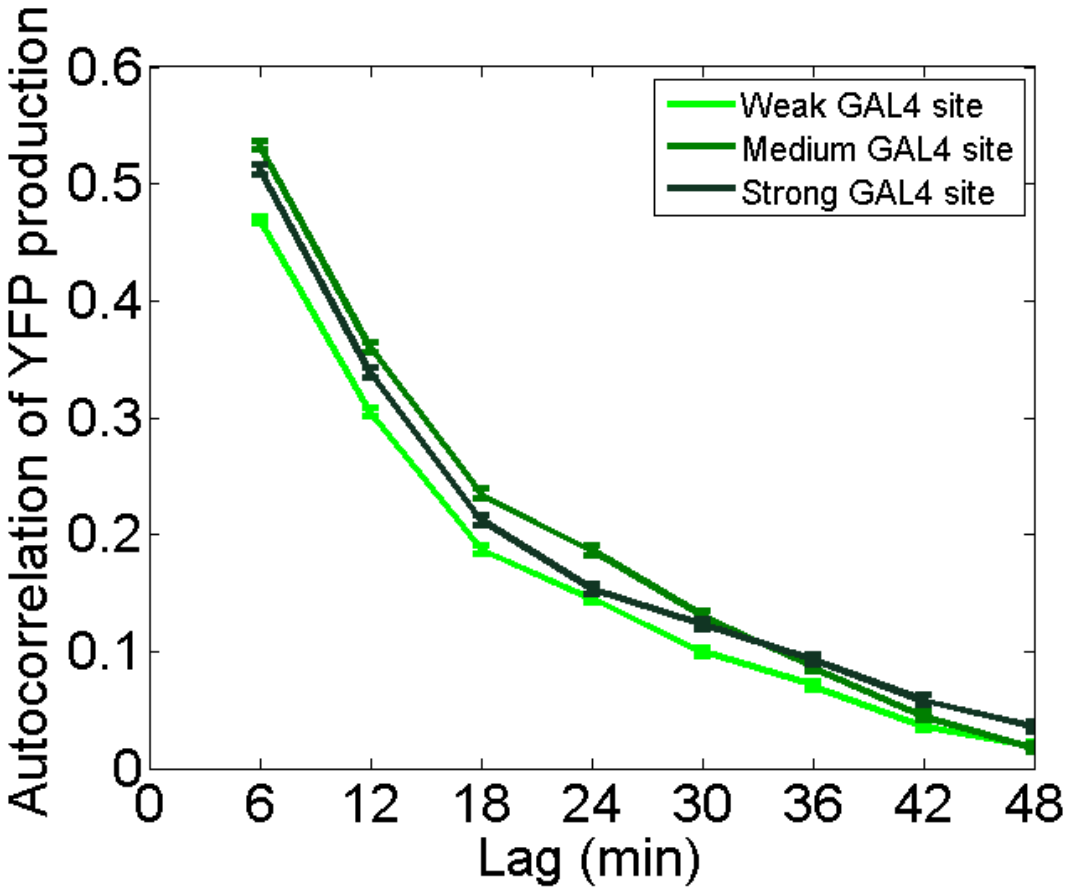

Figure S2. Higher affinity binding sites result in lower promoter dynamics. Shown is the average autocorrelation of normalized YFP production rates across thousands of different cell traces for each of three different promoter variants that differ only in the affinity of the Gal4 site. Bars denote standard errors.

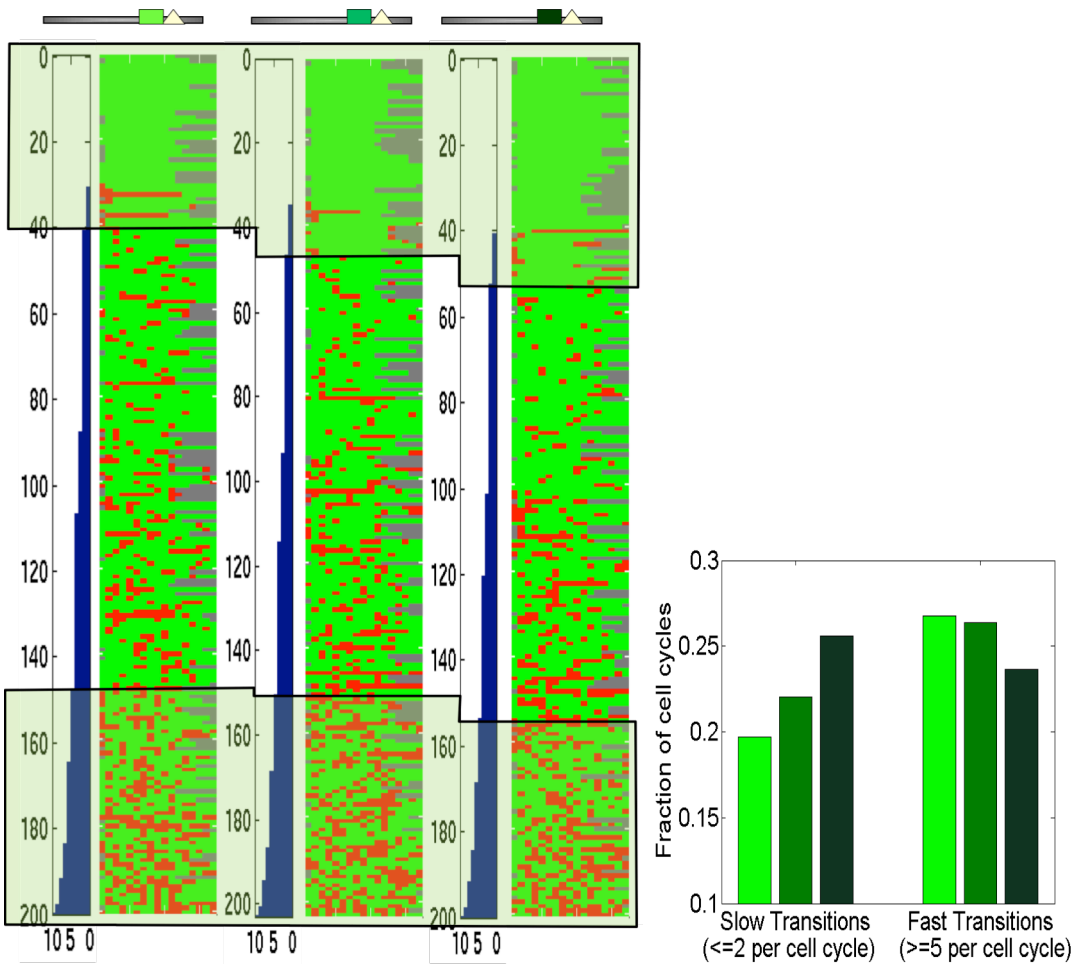

Figure S3. Increasing the affinity of a transcription factor binding site results in a lower rate of transitions between high and low expressing states. Same as Figure 3B and Figure 3D, but for three promoter variants with differing affinities for the Gal4 binding site. The bar graph shows the fraction of cell cycle traces of each variant in which the number of transitions between high and low expressing states was at most 2 (slow transitions, left bar graph) or at least 5 (fast transitions, right bar graph). The comparison of these different promoter variants was done at a threshold in which the fraction of all low expressing states in each variant was 70% (since absolute expression levels vary across variants, the absolute threshold value is different for each variant). The left heatmap shows a visual illustration of the cell cycle traces from the bar graphs where for each promoter variant, shown are 200 rows that each correspond to a time trace of one cell cycle of one cell with colored entries representing high (red) or low expressing states (green) at a threshold in which 70% of all states were low expressing. Rows are sorted according to the number of transitions between high and low expressing states, and the 200 rows were sampled from all cell cycle traces such that they accurately represent the same probability distribution of number of transitions across all cell cycle traces.

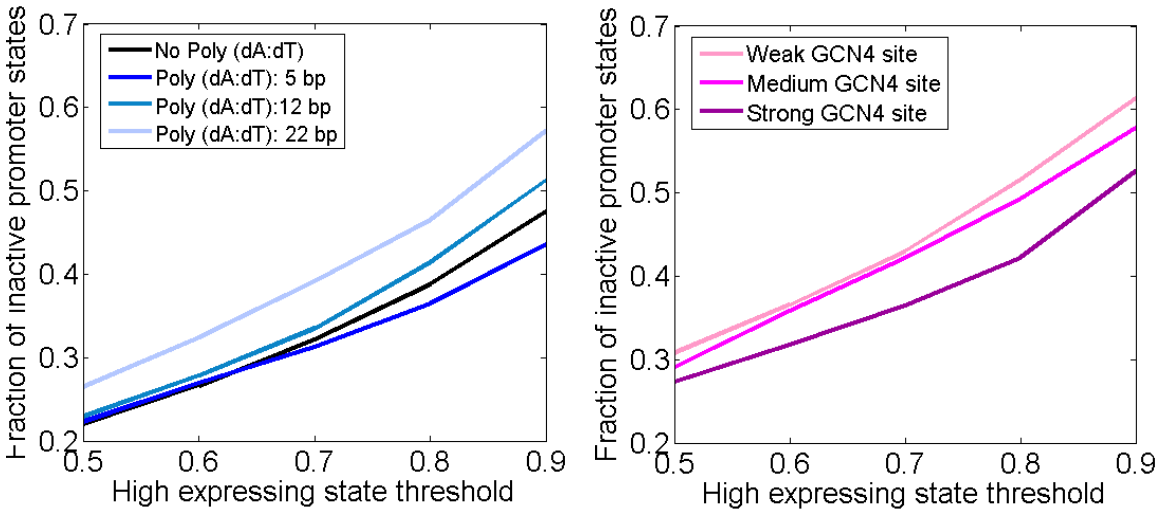

Figure S4. Lengthening poly(dA:dT) tracts and strengthening transcription factor binding sites have opposing effects on the rate of promoter transitions between high and low expressing states. For every promoter variant, we extract each cell cycle of every one of its cells, and classify for each cell cycle the trace of normalized YFP production rate into high and low expressing states according to whether they are above or below a predefined arbitrary threshold, respectively. From these numbers, we can then compute the probability of transitioning between high and low expressing states for every variant. For promoter variants that differ in the length of a poly(dA:dT) tract, the left graph shows the probability of transitioning between high and low expressing states (y-axis) across a broad range of thresholds (x-axis) on the fraction of all low expressing states in every promoter variant. The right graph shows the same computation but for three promoter variants that differ in the affinity of a Gcn4 binding site. Note that across the broad range of thresholds examined, lengthening a poly(dA:dT) tract (left graph) results in a higher rate of transitioning between states, whereas strengthening a transcription factor binding site (right graph) results in a lower transition rate.

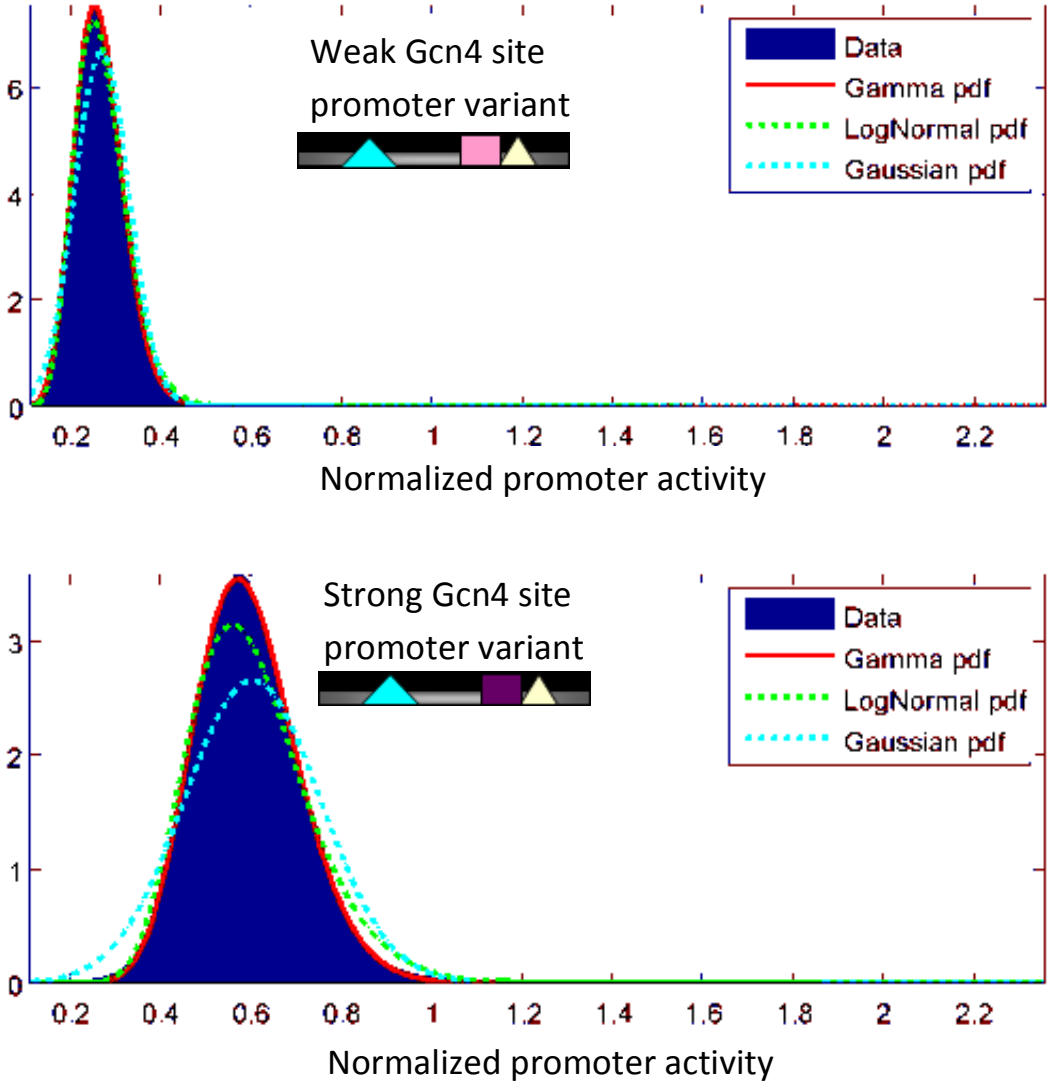

Figure S5. The static distribution of YFP fluorescence accurately fits a Gamma distribution Promoter activity histograms measured by flow cytometry for two representative variants and their relative fits to several distributions: gamma, lognormal and Gaussian. For all promoter variants, the Gamma distribution best fit the data and thus justify the calculations of promoter dynamics.

Supplemental Experimental Procedures

Image acquisition

To ensure monolayer growth of cells and to keep a stable focal plane, cells were imaged in a yeast microfluidic plate (CellASIC). This plate enables constant flow of media throughout the experiment and thus provides “chemostst-like” growth. Images were acquired by a fully automated inverted fluorescence microscope. Raw images were segmented and objects were tracked to yield a single track of fluorescence intensity over time (**Fig. S6**).

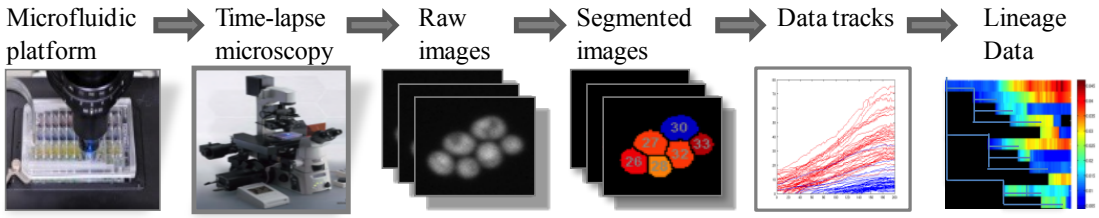

Figure S6. Image analysis pipeline.

Image analysis pipeline

In order to obtain high quality data on the cells imaged in the microscopy system, the images were analyzed in a modular framework (**Fig. S7**), with six modules that deal with **image corrections**, **cell segmentation and tracking** of the cells , **segmentation and tracking post processing** , analysis of the **cell lineage**, **masking and filtering** the data, and **dilution correction and production rates**. The main components of the module framework are the **flat field correction** module which first deals with removing image artifacts, followed by the **background correction** module that removes cell fluorescence that did not originate from the respective cell. The cells are then segmented and tracked using a modified version of **CellProfiler**(Carpenter et al. 2006) . Problems in the segmentation and tracking and several other issues in the image area are then corrected using the **post processing modules**. The **Cell Lineage** module is used to obtain cell lineage assignments for each cell born. The **Masking and filtering** module is used to filter problematic or low quality cell data tracks. Finally, the **dilution correction and production rates** module is used to create tracks corrected for dilution effects, and produce production rates from those tracks.

Figure S7. Modules used in the image analysis pipeline.

1. Image Corrections

1.1 Flat field correction: A flat field correction map is calculated to account for the observed non-uniformity of the images. A set of images was acquired at the beginning of each experiment from a field of view devoid of cells that contains only a homogenous medium (SCD). These images were taken with increasing exposure times of: 0, 10, 50, 100, 200, 400, 800, 1200, 1600, and 2000 milliseconds. Usually, 6 fields of view repeats were done, in which the intensities at these exposure times were measured. A linear regression over the exposure times and their respective pixel intensities was calculated for each pixel, in order to obtain the pixel's gain (the linear regression slope - amount of signal given by the detector as a function of the amount of light) and offset (y axis intersection, auto florescence of medium). This resulted in correction maps: one specifying the gain correction over an entire image and one specifying the offset correction over an entire image. The gain and offset correction maps were then smoothed using a Gaussian filter with a 32 and 16 filter size, respectively, in order to remove noise effects in the flat field images. Finally, a normalized gain map was calculated by dividing the gain map by the average gain over all pixels. Each image in the experiment was then corrected by subtracting the offset of each pixel and dividing by the gain. Any pixel values that were reduced below zero were set to zero.

1.2. Background Correction: After flat field correction, we corrected the images for fluorescence background resulting from auto-fluorescence of both medium and surrounding cells. To remove this fluorescence without removing fluorescence intrinsic to the cell, the fluorescence intensities in proximity to cells were measured and subtracted from the cells. In order to subtract the background, each field corrected image was roughly and widely segmented according to the mCherry to obtain a cell mask (using CellProfiler) with a wide margin around the cells (**Fig. S8**). A subtraction mask was created by dividing each image (mCherry and YFP separately) into 32 by 32 pixels blocks. For blocks with very few cell pixels (at least 85% of the pixels within it occupied by background pixels, pixels not inside the identified cell mask) the median value of the background pixels was assigned as the value of the subtraction mask block. For the rest of the blocks (blocks containing mostly cells), the background was computed according

to the mean of the computed blocks surrounding it. In every case, the next block to be computed was chosen as that with the most computed blocks surrounding it. Thus, blocks containing mostly cells were assigned subtraction values composed of the subtraction values of blocks close to them, since such blocks contain very few background pixels that cannot be extracted reliably. Finally, in order to create a smooth subtraction, the subtraction mask was smoothed using a Gaussian filter of size 32. The subtraction mask was then subtracted from the original image. Any pixel values that were reduced below zero were set to zero. Notably, the background fluorescence intensity represents a rather small fraction (~2%) of the cells' fluorescence intensity.

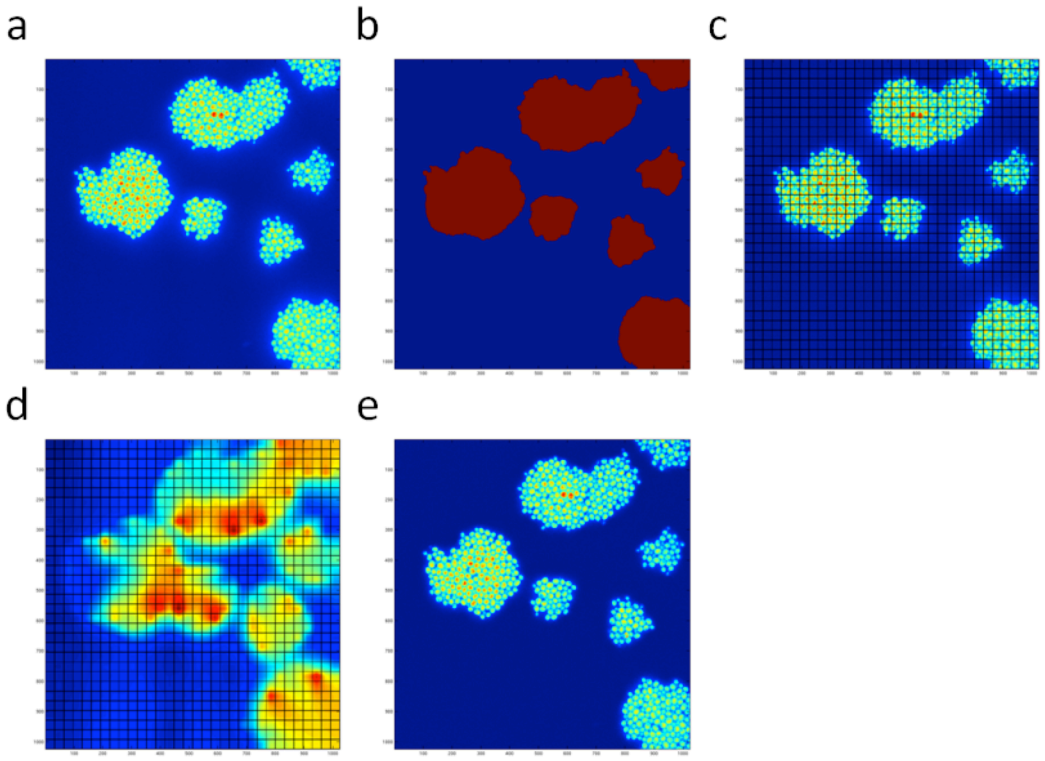

Figure S8. Background fluorescence subtraction. To remove background fluorescence, we used the mCherry flat field corrected images (**a**) and roughly and widely segmented to create a cell mask (**b**). Images were divided into blocks, calculating median background fluorescence in blocks that contain little (less than 15%) cell fluorescence (**c**) and obtaining subtraction values of cell blocks from the values of blocks surrounding them (these subtraction masks were then smoothed) (**d**). The images were subtracted according to the subtraction masks resulting in background corrected images (**e**).

1.3. Bleaching correction: A photobleaching curve was acquired to model the effects of light exposure on the fluorescence trajectories. To obtain a photobleaching curve, a field of view of cells expressing YFP and mCherry were subjected to constant exposure of fluorescent light. Resolution of image acquisition was 1.5 second. The mean fluorescence over a region of interest was plotted against time and this data was fitted to an exponential function to obtain the decay rate, γ_p, for YFP and mCherry, separately, according to Cookson et al(Cookson et al. 2005). We obtained a negligible correction term, of less than 0.015% for YFP and 0.02% for mCherry (the decay rate γ_p equals 0.0015 for YFP and equals 0.002 for mCherry). Therefore, photobleaching correction was not applied to the data.

2. Cell Segmentation and tracking

Image analysis was performed using a modified version of the CellProfiler software, which provides high quality image segmentation and tracking capabilities. Cells were segmented using CellProfiler's automatic adaptive thresholding of the mCherry fluorescent signal, followed by CellProfiler's watershed algorithm to separate clumped cells. Background corrected mCherry fluorescent signal was used for segmentation, as this fluorophore is driven by the same promoter in all of our strains. Modifications to CellProfiler include the handling of misshapen cells (usually as a result of segmenting two cells as a single cell) by iteratively increasing the strength of the mCherry threshold and watersheding attempts on the result. Other modifications were used to reduce computer memory usage. A fixed set of segmentation parameters for all experiments were empirically chosen. Following segmentation, the tracking of cells in CellProfiler was performed by pixel overlap of cells across time. The measured parameters for each cell in every time point were: mean, median and total fluorescence intensity (both for mCherry and YFP), pixel area, cell eccentricity (defined as the ratio of the distance between the foci of the ellipse, with the same second moments of the cell, and the ellipse's major axis length), and x and y centroid locations.

3. Segmentation and tracking post processing

We developed further processing steps to overcome problems that remained with the tracking and segmentation process above. The major modules of the post-processing in their order of operation are:

1. **Trim invasions module** – used to remove data tracks of cells invading from outside the current field of view, as well as to remove data tracks of cells affected by such invasions. This module also filters cells that are improperly tracked.
2. **Split module** – merges cells that are incorrectly split into two cells in the segmentation process, mainly due to vacuoles. An example for a split correction is seen in Figure S9.
3. **Merging module** – corrects data tracks of two cells that incorrectly merge for a short time into a single cell.
4. **Interpolate intensity drops module** – identifies outlier fluorescent intensity values and interpolates them using intensities from neighboring time frames.
5. **Tracking errors correction module** – disconnects cell tracks that are incorrectly merged together in time in the tracking process (or incorrectly merged in the gapping module in the second run of the module).
6. **Clean outliers module** – cleans outliers in data tracks due to segmentation errors based on cell area data tracks.
7. **Gapping module** – connects data tracks of two cells which are actually a single cell with incorrect tracking.
8. Filtration modules – several modules that filter problematic data tracks.
   Remove cells with no neighbors: cells that appeared in some time frame without any neighboring cells next to them in their birth frame were removed (as a cell cannot be born without a nearby parent). This may happen due to poor medium trapping, resulting in cells that drifted far from their previous location in a single frame and therefore were not tracked correctly.

   Remove cells with high birth signal intensity**:** when a bud is formed, it has a relatively lower area and mCherry signal intensity than an adult cell. Cells that first appear with a high birth intensity and large area are usually not born at that frame, but are usually a continuation of poorly tracked cells. Therefore, if a cell had a significantly high total mCherry intensity and a large area at its point of birth compared to the rest of the cells at their birth frame (Z-Score of total mCherry birth intensity higher than 2 and Z-Score of cell area at birth higher than 2), its data tracks were completely filtered.

   Remove cells on edges**:** cells in close proximity to the edge of the field of view may produce poor segmentation and incorrect signal as usually part of their

volume is outside of the field of view. Therefore, the part of the data tracks in which a cell appears too close to an edge (defined by a 14 pixel margin) was removed to avoid incorrect data tracking in those frames.

Remove cells that appear in too few frames**:** Cells appearing in only one or two frames were completely filtered, as these cells are usually leftovers of other cell tracks that were not combined with these tracks.

4. Cell Lineage

One of the prominent advantages in dealing with single cell data is the ability to track cell lineages. This information was obtained by discovering the most likely parent cell assignment for each new bud in each field of view using the Cell Lineage algorithm. The basic idea of the algorithm (**Fig. S9**) is to compute scores for each of the possible parents, and then assign the best scoring parent for each of the buds, taking into account the effect of the assignment on the scoring of the other bud-parent pairs. This is done by optimizing for the best scores over many parent-bud pairs.

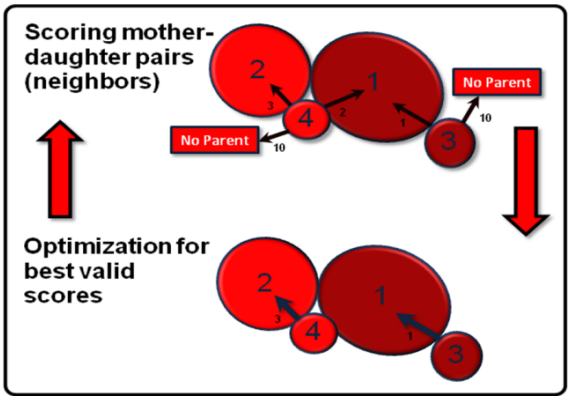

Figure S9. Parent assignment scheme. Scores are computed for each mother daughter pair. These scores are optimized to obtain the best pairs. These pairs are later used to create better scores.

Computing the scores: For each bud, a score was calculated for each of its potential parents (its neighboring cells at the time of its birth). The scores were based on the cell cycle of the parent cell. As a bud emerges, it is tracked separately from the parent cell, and thus, the fluorescent protein produced by the parent decreases due to diffusion to the emerging bud. This happens from the synthesis stage, through the G2 phase until mitosis, when the parent and bud disconnect. At the start of the next G1 phase, the

parent cell again accumulates new fluorescent proteins. This scheme was illustrated well by Cookson et al. (Cookson et al. 2005) and is shown in Figure S10 over real mCherry total intensity. The S phase (synthesis) and the M phase (mitosis) were used as reference points to define the cell cycle of a parent. These two points were identified around the birth of each bud for each of its possible neighbors at the time of its birth.

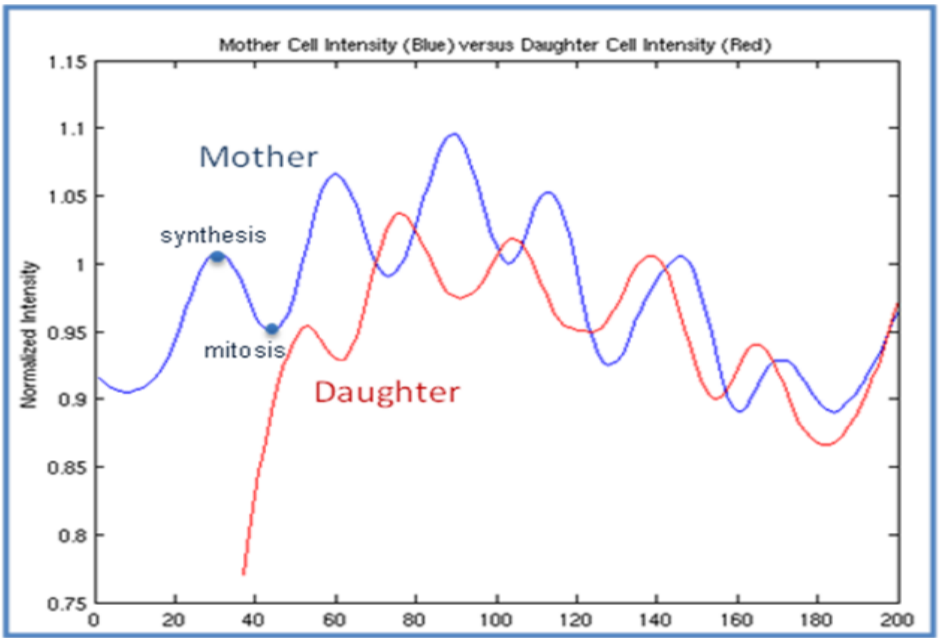

Figure S10. Cell cycle intensity fluctuations. Shown are mother and daughter cells and their total mCherry signal. At the start of the mother's cell cycle, the signal increases. From the S phase (synthesis) to the M phase (mitosis), mCherry diffuses towards the daughter cell.

The behavior of these neighbors around these points helped to create scores defining how likely these neighbors were as the parent of the emerging bud. In order to identify the possible cell cycle points in each neighbor, the total mCherry signal was modified to better display the cell cycle trends. Outliers in the signal were smoothed as depicted in the post processing (in the Clean Outliers module (3.6)) but using a smaller moving window of 5 frames. The signal was then smoothed using a moving window of 5 frames. Next, the period in which the cell itself buds from its parent was removed (filtering until the point in which the cell's area Z-Score at the first 150 minutes of life rises above -1), as in this period a cell cannot produce a bud, and the swift rise in total intensity at that period would have strongly influenced the global smoothing described next. The general

signal trend of the cell was then obtained using a strong local regression smoothing (loess smoothing). This general signal was subtracted from the signal to produce a signal which was not affected by the general trends of the cell, but mostly by the cell cycle. To make the signal comparable between cells, the result was then scaled according to the signal's mean and standard deviation. This signal was then used to find local maxima peaks around the birth of a bud to find the S phases (54 minutes prior to the birth and up to 27 minutes after the birth) and minima peaks to find the M phases (23 minutes prior to the birth and up to 70 minutes after the birth). Note that due to segmentation issues, it may happen rarely that the M phase occurs before the birth, or the S phase occurs after the birth, but an S phase occurring prior to the M phase was not allowed. Also, a 36 minutes minimum time was set from the cell's birth until its first S phase.

Several parameters were used to create the score for each of the possible parents:

- Time from S phase to bud birth.
- Time from bud birth to M phase.
- S phase peak shape – Scaled signal values of the potential parent from 20 minutes prior to the S phase peak until 20 minutes after the S phase peak.
- Eccentricity of potential parent cell around the S phase (at 20 minutes prior to the S phase peak until 20 minutes after the S phase peak), as the shape of a parent cell is somewhat eccentric close to the budding point.
- Area of potential parent around the S phase (at 20 minutes prior to the S phase peak until 20 minutes after the S phase peak), as a parent cell slightly expands prior to the S phase and contracts slightly after the S phase.
- M phase peak shape – Scaled signal values potential parent from 20 minutes prior to the M phase peak until 20 minutes after the M phase peak.
- Neighboring time - Fraction of time bud and potential parent neighbor each other (less than 5 pixels distance between cell edges) from bud birth to M phase (as the parent and bud are attached until the M phase, this number should be very close to 1, and therefore a neighboring fraction less than 0.4 was not accepted).
- Eccentricity of potential parent from 20 minutes before the bud birth until its birth.
- Ratio of the area between the potential parent and the bud from 15 minutes prior to the M phase up to the M phase. There is a certain ratio between the parent and the bud as the bud grows which this property tries to capture.

- Total mCherry intensity ratio (outlier cleaned and smoothed) between the potential parent and the bud from 15 minutes prior to the M phase up to the M phase. This ratio should increase in time as the protein diffuses into the bud, resulting in a decrease in the intensity of the parent and an increase in the intensity of the bud.

To use the above parameters to identify the real parent out of the neighboring cells, we sought to use a Naïve Bayes classifier, which models the distribution of these parameters for real parents and non-parent neighboring cells. The score was defined as:

$$Score = \frac{P(M=-|X)}{P(M=+|X)} = \frac{P(X|M=-)}{P(X|M=+)} \cdot \underbrace{\frac{P(M=-)}{P(M=+)}}_{constant}$$

$$\propto \frac{P(X|M=-)}{P(X|M=+)} = \prod_{p \in parameters} \frac{P(x_p|M=-)}{P(x_p|M=+)}$$

Definitions:

False parent	M = -
True parent	M = +

$$P(x_p|M=m) = \sqrt[n]{\prod_{i=1}^{n} P(x_{p,i}|M=m)} \; where \; P(x_{p,i}|M=m) \sim N(\mu_{p,i,m}, \sigma_{p,i,m})$$

The probability distribution of all parameters was taken to be a Normal distribution. Some parameters were computed from multiple time points (such as the S phase peak shape defined across 40 minutes), by geometrically averaging the probabilities over all time points. The distribution in each point was modeled over averaged values (median and standard deviation over 75% of the values around the median).

As the positive set, we took parent-bud pairs whose bud only had one neighbor during their birth (and thus this neighbor must be the parent). In addition, we required that neighbor to be a neighbor for at least 80% of the period between the bud's birth and the M phase, and we required that it was not weaker in intensity than the bud at any point during that time. In the choice of the negative set, it is important not to enter true parents into it. Therefore, only neighbors of buds with 4 or more neighbors were considered into the negative set. The best scoring neighbor out of those was not entered into the negative set, considered as the true parent, and the rest were entered. The probability of each neighbor to be a parent was then calculated using the positive set distribution (score calculated only as: $P(X|M=+)$). Thus, the probability of entering a true parent

into the negative set in each choice was at most 0.25 (if we consider the classifier according to the positive set to be at least as good as a random classifier).

Several penalties were used to penalize empirically unlikely events:

- If the neighboring ratio parameter was lower than 0.6, 2 standard deviations were added to the score (standard deviations according to the middle 90% of the scores).
- If the distance from birth to M phase parameter had a Z-Score greater than 3, a single standard deviation was added to the score.
- If the distance from S phase to birth parameter had a Z-Score greater than 3, a single standard deviation was added to the score.

In order to decrease the search space for possible parents, parents that had a Z-Score (built from all the scores) above 4 were filtered. Also, only up to 4 possible parents were allowed per bud (the best scoring ones were kept). A score for a no-parent assignment was set as a score having a Z-Score above 5. That way, only when there was no other choice, no parent was assigned.

Score optimization: Since the optimal score does not always reflect the correct assignment, a certain optimization on the scores was needed, instead of greedily picking the best scoring parent for each bud, as depicted in figure S11. The optimization was not global across the entire movie, since this was not computationally intractable. Therefore, a local optimization was done for each new bud. The optimization considered parent to bud assignment scores, and also other nearby bud scores that might be affected by the parent assignment (up to a certain degree). This way, for each bud born, we found the best parent assignment, which also did not hinder the parent assignment of the buds that could have been affected by its assignment. For each new bud (denoted as the center bud), buds affected by it were chosen as buds born up to an hour after its birth and in layers of influence around it. Layers of influence refer to layers of bud assignments around the bud that might be affected by the center bud. The first layer consisted of buds that shared the possible parents of the center bud. The second layer consisted of buds sharing the possible parents of buds in the first layer, and so on, up to 5 layers of influence. This resulted is a search space consisting of all the possible parent assignments for each of the buds affected up to 5 layers from the center bud.

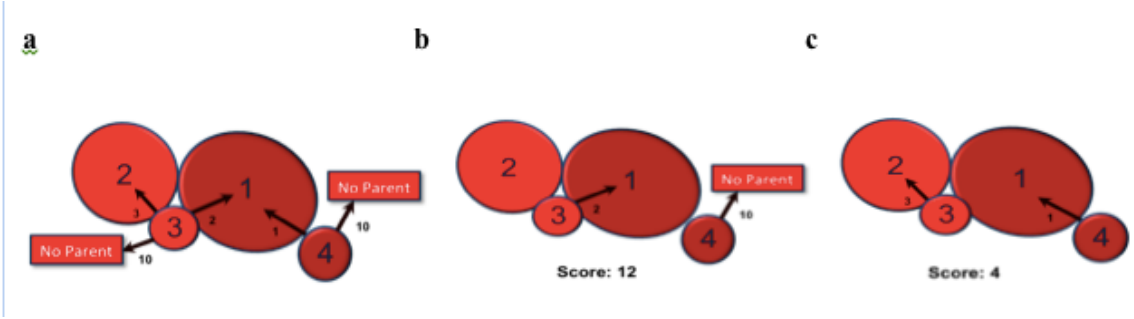

Figure S11. Parent assignment possibilities. An example of possible parent cell assignments. (a) Scores for each of the possible parent assignments for cells 3 and 4 (the lower the score, the better). (b) A simple greedy assignments – assigning best possible parent for cell 3, and then the best possible parent for cell 4 (cell number 1 is already assigned to cell 3, and as a cell can only have one bud at a time, cell 4 is assigned no parent), resulting in a score of 12. (c) Optimization on the scores, to assign the best possible scores for each bud, assigning cell number 2 as the parent of cell number 3, and cell number 1 as the parent of cell number 4, resulting in a score of 4.

A valid assignment was defined as an assignment in which no two buds were assigned the same parent at the same time. The optimal valid assignment was found for the entire search space, and only the center bud was assigned a parent. After assignments were made, the scores were calculated again, but with the positive set taken as the assigned parents, and the negative set as the incorrect parent of those same buds.

In order to further improve the scores, the distribution of the time between cell cycles and time from the birth of a cell until its first cell cycle were collected from the previous assignments. Empirical probability density functions were created for both measures for the positive set. For the negative set, a uniform density function was created as these measures are meaningless for a negative set. Thus, in the second assignment iteration, the scores for parents budding for the first time were multiplied by a constant divided by the empirical probability density function collected previously. The same was done for the time between cell cycles according to time distance to assignments already made. This algorithm was tested on ~200 manually tagged mother-daughter pairs, and had an accuracy of 0.96.

5. Masking and Filtering

A conservative framework for filtering the data was created to ensure that only high quality data is kept in the end. The following are the standard filtrations:

Mask last known cell cycles: As cell tracking ends abruptly due to the end of an experiment, the last cell cycle of all cells is incomplete. For the majority of our analyses, we required complete cell cycles and therefore the last cell cycle of every cell was masked.

Mask initial growth: The initial growth stage of a bud while it is attached to its parent was usually of no interest to us, as we attributed the protein production at that time to the parent cell. Therefore, the time a cell spent as a bud was masked.

Mask abnormally large and strong cells: Rarely, there are cells that grow to abnormal sizes. These cells usually begin their life as normal cells, and at some point in time grow to large proportions. To avoid such extremities in the analyses, these cells were filtered from the point of their abnormal growth. These points were found by finding cells with an area Z-Score larger than 2 and a total fluorescence intensity Z-Score larger than 5 for more than 5 consecutive frames.

Mask missed cell cycles: As some cell cycles are not detected by the lineage, the cell cycles that follow appear as very long cell cycles. Usually during these incorrectly long cell cycles, the missing S phase is detectable. Within these cell cycles the probability for an additional S phase peak was calculated. If an additional S phase was found with high probability, the two cell cycles were filtered.

Long cell cycles were found by finding cell cycles longer than 1.5 times the median cell cycle length in the population. Cell cycles longer than 2 times of the median cell cycle length were automatically filtered as being too long. The cell cycles in between these lengths were checked for an additional S phase peak. The probability for an S phase peak was built from several parameters used in the Lineage algorithm:

- S phase peak shape
- Eccentricity of cell around the S phase
- Area of cell around the S phase

The probabilities according to these three parameters were sampled from S phases set by the Lineage algorithm. The median and standard deviations were taken from the top 80% sample in order to consider only high probability S phases, which are more likely to be true. Then, the threshold for S phase peak was set as the median minus two standard deviations. Long cell cycles were filtered in cases in which the probability for

the additional S phase to occur passed the threshold, within its expected time frame in the cell cycle. The expected period for the additional S phase was set from the time of the previous M phase plus the minimum empirically checked G1 time and until the time of the S phase of the cell cycle minus the minimum empirically checked interphase time.

Mask cell cycle arrest: Due to high cell density within a field of view many cells enter cell arrest. These occurrences were found and filtered by searching for cell cycles longer than 3 times the median cell cycle length in the population. In a cell arrested field of view, it is probable that cells not in a cell arrest also suffer in some way from the conditions in the field of view. Therefore, in field of views in which more than half of the cells were in a cell arrest, all the cell cycles from that point onwards were filtered.

Mask negative protein production cell cycles: Another strong indication of missed cell cycles are cell cycles that have a long period of negative mCherry production (see no dilution section). As the mCherry has very slow degradation time, a negative protein production can only happen due to protein dilution to the bud. This dilution should be accounted for in the calculation of the non-diluted total protein fluorescence, and the primary reason for such a drop, is a missed cell cycle. Protein dilution occurs during the bud formation between the S phase and M phase, and therefore the typical drop length was calculated as the median length of time between the two phases. If within 33% of a window of such size negative production occurred, the cell cycle in which the drop occurred was filtered.

Mask low quality cell tracks: During the post processing framework some cells have more problems than most, e.g., cells with large vacuoles might split many times. The quality of these cell tracks is questionable as the post processing framework cannot always find a perfect solution. Such cells, exceeding certain thresholds (defined below) on the amount of problems that occur in the post processing, were considered as low quality cells and filtered from further analysis. For a cell to be considered as a low quality cell, a cell needed to pass any of the following criteria:

- Split at least three times (found in the Splitting module).
- Disappear and reappear at least three times (in the Gapping module).
- Having at least 3 long intensity drops (of at least 3 frames).
- A tracking error of at least 3 frames length occurred.

Mask growth rate: In order to achieve relatively constant conditions, we aimed at keeping the growth rate at certain limits within a strain, filtering periods in field of views where the median growth rate differed from a reference growth rate obtained from all the

field of views. We defined the growth rate at a certain point in time as the median of the cell cycle times at that point in time. To obtain a reference growth rate, we calculated the growth rate for each field of view at a time reference between 150 and 420 minutes into the experiment, which is usually a time of stable growth and low cell density. The reference growth rate was set as the median over the median growth rate over time of each field of view. The limit for field of view growth rate was set as up to 15% distant from the reference growth rate. Only time periods within a field of view in which the growth rate was within the limits for a time period were kept. The length of these periods needed to be at least twice the reference growth rate so as to keep only large periods of almost constant growth rates. For periods shorter than 4 times the reference growth rate, between periods of growth rates within the limits, slight deviations of up to 22.5% from the reference growth rates were allowed. Note that in order to complement this work in terms of comparisons between repeats, different strains and different experiments; we also manually checked that the resulting growth rates were similar.

6. Dilution correction and production rates

In order to perform analysis on protein production rates, we defined the protein production rate as the average amount of protein accumulated over a certain period of time. However, the total protein fluorescence of a cell provides information only on the protein accumulated in a cell without accounting for the protein and transcripts lost through dilution to its buds (mCherry and YFP degradation are negligible in relation to dilution. This was verified by observing negligible degradation in mCherry and YFP after addition of a translation inhibitor to the media). As the bud transcripts and cell environment are transported from the parent through active transportation and diffusion, we considered the total protein accumulated in the bud until the mitosis separation as proteins produced by the parent. Although the bud is a part of the parent until the mitosis point, it is considered as a new cell by the image segmentation and tracking. Therefore, the protein accumulated in the bud until the mitosis point was added to the protein accumulated in the parent as depicted in Figure S12, to obtain the protein accumulated in the parent until the mitosis point while correcting for dilution to the bud. The corrected accumulated protein was then smoothed for outlier points in the signal (as done in the post processing, in the 'Clean outliers module', but with a smoothing window of size 5).

After calculating the corrected accumulated protein in a cell for each time point, the calculation of the protein production rate from time T to time T+ΔT for a single cell was defined as:

$$Production\ rate\ (T \ldots T + \Delta T) = \frac{Accumelated\ protein(T+\Delta T) - Accumelated\ protein(T)}{\Delta T}$$

Prior to the production rate calculation, the accumulated protein signal was smoothed separately in each of the time windows in which the protein production rate was calculated using Matlab's Loess smoothing, with a span set to 0.4 of the data points in the window, such that data from a certain time window did not leak to its adjacent windows.

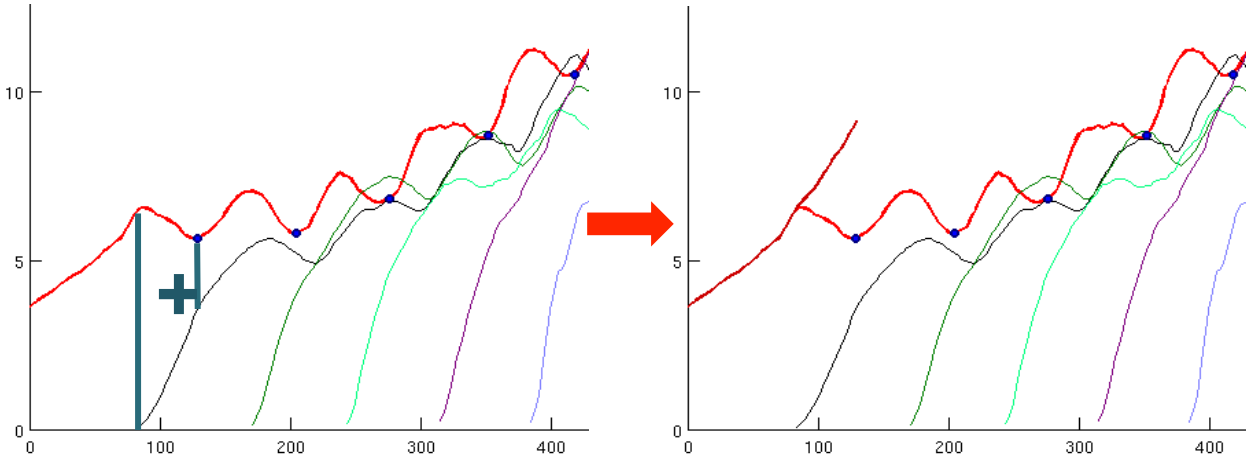

Figure S12. Correcting for dilution. An example of how the effects of dilution are corrected. The bud total intensity values (dark blue) are added to the total intensity of the parent (red) from the point of the bud's appearance until the mitosis point (first blue dot) to produce the parent's total intensity values, corrected for dilution (dark red). The rest of the lines are additional daughter cells, and are also used for dilution correction.

Yeast Strains

A shared master strain that contains the YFP reporter gene, the HIS3 proximal promoter (100bp upstream of the ATG), and the mCherry fluorescent protein downstream of the *TEF2* promoter, was used to construct the library. The variable part of each promoter were synthesized by Biomatik and inserted into a master strain by genomic integration, as previously described (see (Raveh-Sadka et al. 2012) for details).Additional strains with mutated Gal4 binding sites were designed to span a range of weak medium and strong affinity to Gal4, based on previous studies by Ptashne et al. as follows:

1. STRONG: CGGAAGACTCTCCTCCG (“near consensus”, (Ginigea and Ptashne 1988))
2. MEDIUM: AGGAAGACTCTCCTCCG (GAL1-GAL10 UASg site3, (Liang et al. 1996))
3. WEAK:CGGATTAGAAGCCGCCG (GAL1-GAL10 UASg site1, (Liang et al. 1996))

Flow Cytometry data analysis and Gamma distribution analysis

An automatic data analysis pipeline was applied for gating and filtering the data and to remove outliers. Wells were considered outliers having abnormal forward scatter, side scatter or mCherry distributions. Cells collected in the first or last 0.5 seconds in each well were discarded from further analysis, as well as cells for which negative or saturated values were measured in one of the parameters. Cells collected over periods that show flow instability (bubbles, etc.) were also removed. Automated gating based on FSC and SSC values was implemented to minimize cell heterogeneity and reduce extrinsic variability. The gating algorithm select for the lower population that is enriched for G1 cells. The final gated YFP values were normalized by mCherry.

Fluorescent proteins

A direct and accurate measure of promoter activity over time was achieved by real-time monitoring the fluorescent protein levels as a measure of promoter activity. The YFP protein was very stable over the acquisition time (>15 hours), showing no degradation, also after addition of cycloheximide to block translation. Protein accumulation, corrected for dilution from mother to daughter cells, showed a constant increase (Fig. S13) and was considered as an accurate measure for promoter activity.

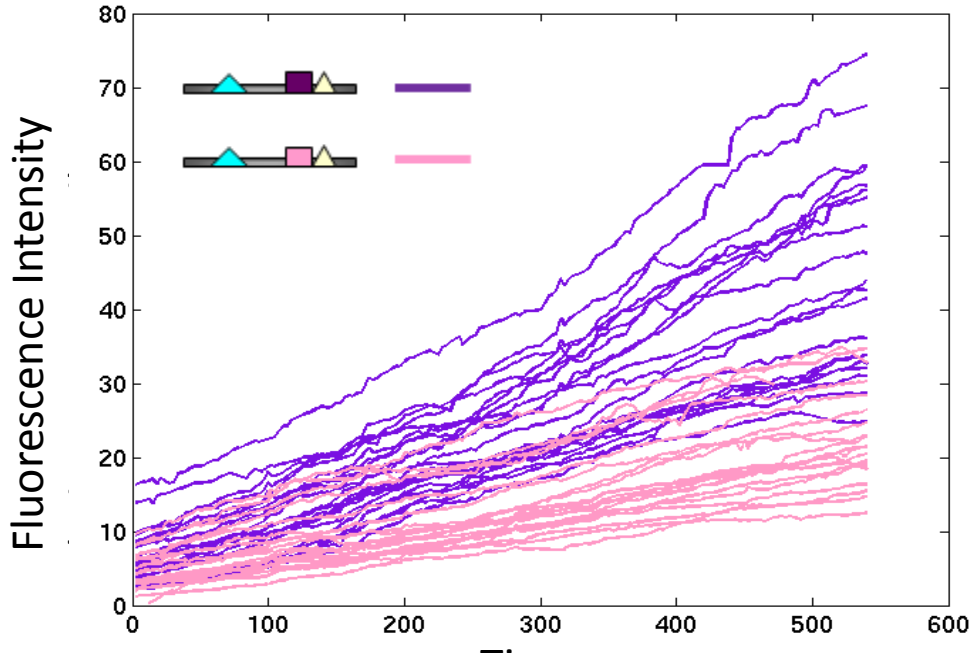

Figure S13. Individual cell tracks, corrected for dilution between the mother and the daughter cells. Cell tracks are shown for 2 strains: strong binding site affinity (purple) and weak binding site affinity (pink). The tracks show robust continuous increase in signal intensity, and were differentiated to yield the production rate (as described in the main text).

Normalization between experiments and wells

As the mCherry distribution should be equal across all wells in all experiments, we made sure that all mCherry rate distributions have the same mean mCherry rate. This was done by measuring the mean mCherry rate from a reference well, and multiplying all mCherry rates of a well by a factor such that its new mean mCherry rate will have the same mean mCherry rate as the reference well. A multiplication factor between the different mean mCherry rate distributions reflects a global factor that should also be reflected in YFP changes as well. Therefore, the YFP rates were also multiplied by the same factor multiplying the mCherry. Displayed in figure S14 are corrected mCherry rate histograms and their common distribution (a distribution created from an equal amount of data points taken from each strain). It can be seen that the mCherry rate histograms of different strains are very similar, and fit the common distribution.

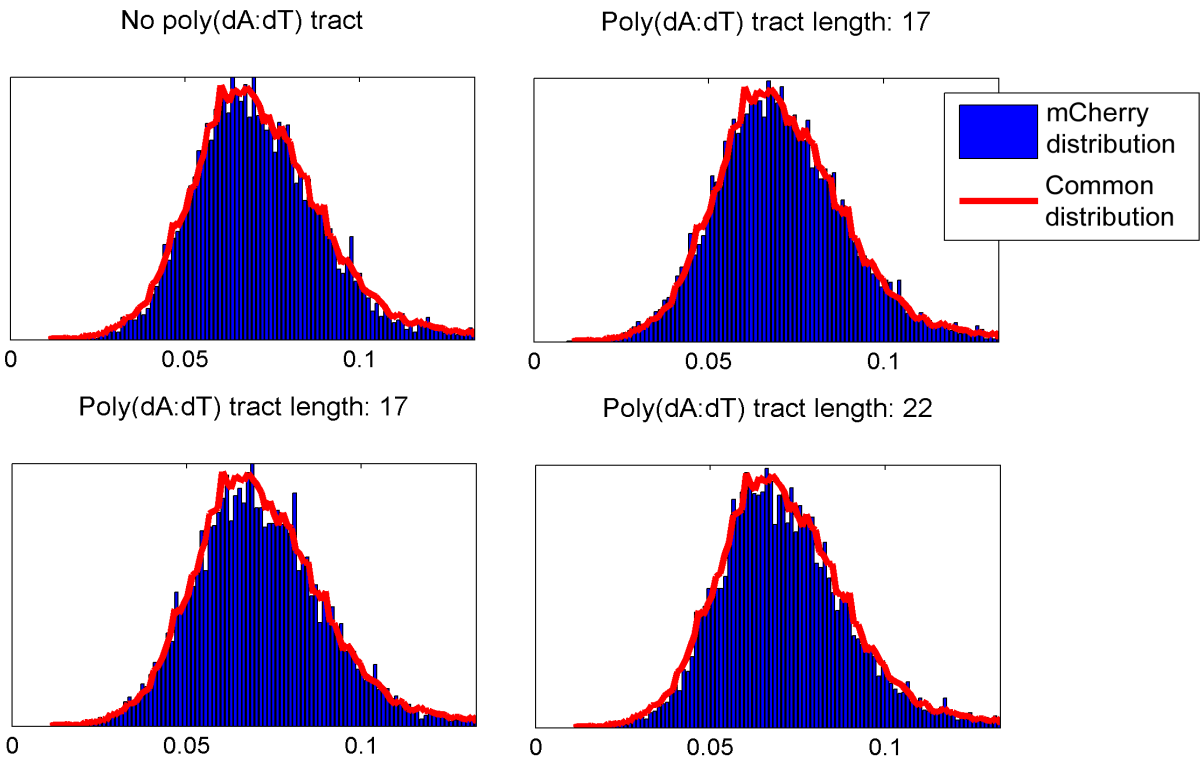

Figure S14. Corrected mCherry rate histograms. Representative mCherry rate histograms normalized between wells and experiments. Depicted in red is a distribution of mCherry rates of all strains in the example.

Autocorrelation of promoter production rates

Normalized YFP rates were calculated for each cell at a resolution of 3 minutes. Negative rates were filtered. The top and bottom 0.5% values of the data for both mCherry and YFP were considered as outliers. Cell cycles containing at least 50% unfiltered YFP and mCherry rates were used. A cell cycle not containing at least 30 minutes of data was not used. Autocorrelations were calculated at a resolution of 6 minutes, averaging the temporal information to reduce noise. Autocorrelation at 3 minutes resolution (an example shown in Figure S15) yielded lower autocorrelation values, however the relation between the different strains was kept the same. Autocorrelation lag information smaller than 14 data points was not used. Then, all autocorrelations across all cells were averaged. Correlations between consecutive time windows were calculated by averaging the values across each resolution (e.g. for a resolution of 9 minutes the values of windows i, i+1, i+2 were averaged), and then computing the correlation of the consecutive windows in a single cell. Then, all the values across all cells for each resolution were averaged. Values filtered were not used for the correlation. Correlations with less than 28 data points divided by the current resolution (where a resolution of 3 minutes is considered as 1 as this is the system's minimal resolution), or less than 4 data points, were not used. To verify that the strains differ in their autocorrelation of YFP rates and not the mCherry rates, we calculated the autocorrelation also for YFP and mCherry rates alone. We clearly observed that for all the strains there was no significant change for between the autocorrelation curves of the mCherry, as opposed to the autocorrelation curves of the YFP rates.

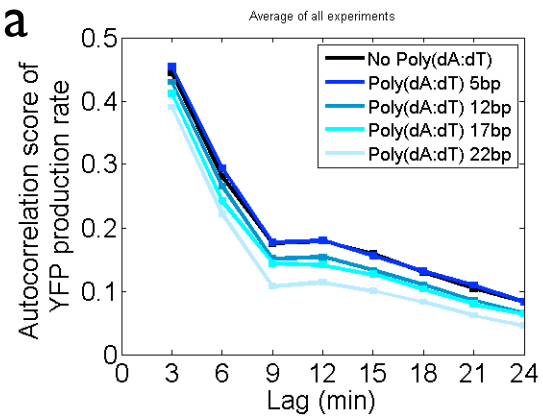

Figure S15. A utocorrelation of normalized YFP production rates at 3 min resolution. The autocorrelation was calculated across thousands of different cell traces for each promoter variants. Bars denote standard errors.

Model of stochastic gene expression

We use a two-state kinetic scheme in which the promoter switches between an active and in-active promoter state, and has transcription, translation, mRNA degradation and protein degradation (Fig. S16).

Figure S16: Kinetic model of gene expression. The scheme represents a promoter that switches between a transcriptionally inactive and active state. On- and off switching happens with rate *Kon* and *Koff* respectively. Transcription, translation, mRNA and protein degradation rates are not shown, but are taken into account in all simulations.

To analyze the temporal dynamics of gene expression we use a stochastic simulation of this model (Gillesple 1977). For steady-state gene expression and cell-to-cell variability (noise) we use an analytical solution of this same model (Sanchez et al. 2011), which was solved using the master equation.

For both models we use the following rate parameters: (All rate parameters are in $minute^{-1}$)

1. **Protein degradation**. YFP was found to be highly stable (see fluorescent protein paragraph). We therefore assume that protein degradation comes only from dilution. Since we measured the doubling time of our strain (~90 min), we chose to fix the protein degradation rate. We set the degradation rate to ln(2)/90 = 0.0077.
2. **Translation rate.** Yeast, under fast growing conditions, has a protein production rate of between 6500 and 19500 proteins/cell/sec (von der Haar 2008). Yeast has around 60,000 mRNAs/cell (Zenklusen et al. 2011). This gives a translation rate between 6.5 and 19.5 proteins/mRNA/min. We set the translation rate to 10 proteins/min.
3. **Transcription rate.** The transcription rate is the rate of production of stable mRNAs while the promoter is in each respective state (on or off). Experimentally

measured transcription rates combine both on and off promoter states. Therefore, the total rate of transcription for a gene is the sum, for all states, of the fraction of time spent in that state times the transcription rate in that state. The upper bound for expression rate is between 4 transcripts/minute and 10 transcripts/minute (Pelechano et al. 2010). We note that these rates represent the combination of on and off promoter states. We set the transcription rate to 5.

4. **mRNA degradation.** The median mRNA half-life was measured to be ~20 min (Wang et al. 2002). We therefore set the mRNA degradation rate to ln(2)/20.
5. **Promoter switching rates (*Kon*, *Koff*)**. Promoter switching is a function of TF binding and unbinding, however each (un)binding event does not necessarily switch the promoter to another state. For this reason promoter switching is much slower than TF binding kinetics. Promoter switching was measured to be in the range of 1e-3 to 1e1 (Zenklusen et al. 2011; Octavio et al. 2009; Tan and van Oudenaarden 2010). For the "normal" promoter we use 2^{-3} min^{-1} and 2^{-2} min^{-1} for *Kon* and *Koff* respectively. For the "fast" promoter we use 1^{-2} min^{-1} and 2^{-2} min^{-1}, for the "slow" promoter we use 2^{-3} min^{-1} and 2^{-3} min^{-1} for *Kon* and *Koff* respectively.

Stochastic simulation of fast and slow promoter dynamics

We simulate three regimes: **normal**, **slow** and **fast** dynamics. The "fast" regime has a higher *Kon* to simulate an increased accessibility with the reduction of nucleosome coverage as a result of the addition of a poly(dA:dT) element. The "slow" regime has a lower *Koff* to simulate the increase in transcription factor binding affinity. **Figure 1** shows three example runs in which the promoter switches between the ON and OFF state, and transcription and translation occur when the promoter is ON. The protein production shows the bursting behavior of gene expression. The figure illustrates that protein production is correlated with promoter dynamics, however also that protein production has a delay compared to promoter on-switching and that, due to chance, not every on state results in a protein production burst.

Autocorrelation analysis of simulated protein production

We found that, in experiment, increasing the binding affinity or decreasing the poly(dA:dT) length both increase the autocorrelation (Fig. 2 from main text). To investigate if this change in autocorrelation is expected given the simple model of promoter state switching, we performed the same autocorrelation analysis on simulated data obtained from the Gillespie simulations. We model the increase in binding affinity by a decrease in *Koff* and the decrease in polyT length by a decrease in *Kon.* Figure S17 shows that indeed we observe the same increase in autocorrelation when we decrease *Koff* (increase affinity) or decrease *Kon* (decrease polyT length).

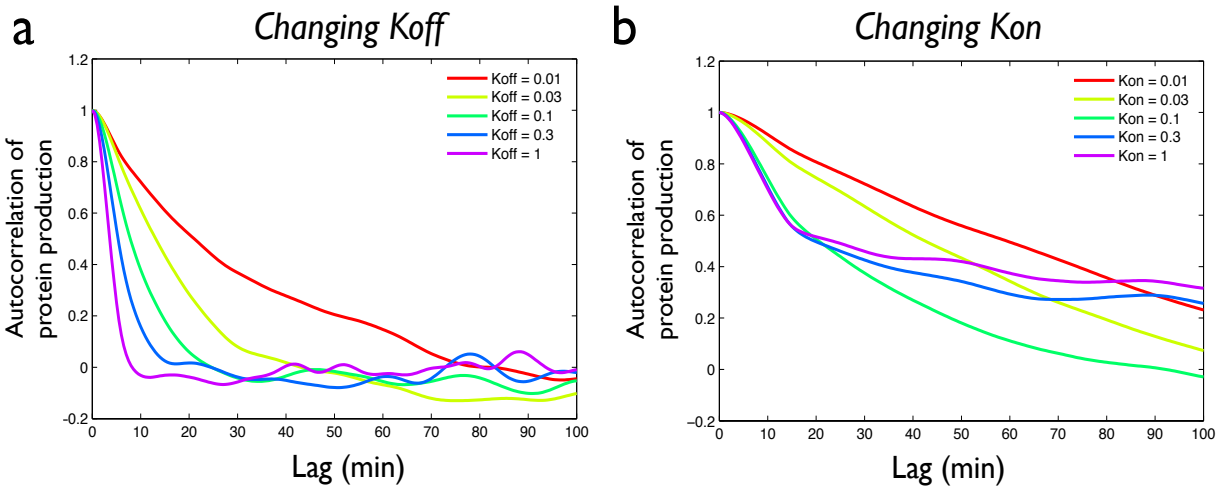

Figure S17: Autocorrelation analysis on simulated protein production. We use the stochastic simulation data to look for autocorrelation changes when changing either *Koff* (a) or *Kon* (b). We observe that decreasing either *Koff* or *Kon* significantly increases the autocorrelation, in line with our experimental observations. Values shown for *Koff* and *Kon* are in min^{-1}.

Limits of measuring promoter dynamics through protein production

In order to systematically investigate how promoter dynamics relate to protein production bursting we use the above stochastic simulation and gradually increase either *Kon* or *Koff* and quantify protein production bursting. We quantify the burst frequency as P(on|off) and burst size as $P(off|on)^{-1}$. $P(on|off) = N_{off->on} / t_{off}$ and $P(off|on) = N_{on->off} / t_{on}$. Where $N_{off->on}$ is the number of *off* to *on* event and $N_{on->off}$ vice versa. t_{off} and t_{on} is the total time in the *off* or *on* state respectively.

To simulate our microscopy experiments as best as possible we sample the protein level every 6 min. We note however that our simulation does not capture every process that affects the delay between on-switching, transcription and translation, such as mRNA maturation and protein folding. However, adding such delays would not change the qualitative result of our simulations.

Figure S18 shows how changing promoter ON and OFF switching (*Kon* and *Koff* respectively) affects mean expression and the derived promoter dynamics from protein production bursting. The figure shows that there is a regime in which promoter *Kon* and protein burst frequency change linearly with respect to each other. However, at higher *Kon* the protein burst frequency saturates. For *Koff* the same happens, but now both at very low and very high *Koff.* There appears to be only a limited regime where changes in *Koff* are linear to changes in the observed protein burst size. However, in the biologically meaningful regimes of *Kon* and *Koff*, the simulation shows that changes in *Kon* and *Koff* can be detected, at least qualitatively, by measuring changes in protein production bursting. The fact that there exists a regime in which this measurement is non-linear could explain why, in our microscopy experiment, we measure a smaller change in bursting than is expected from the change in mean expression.

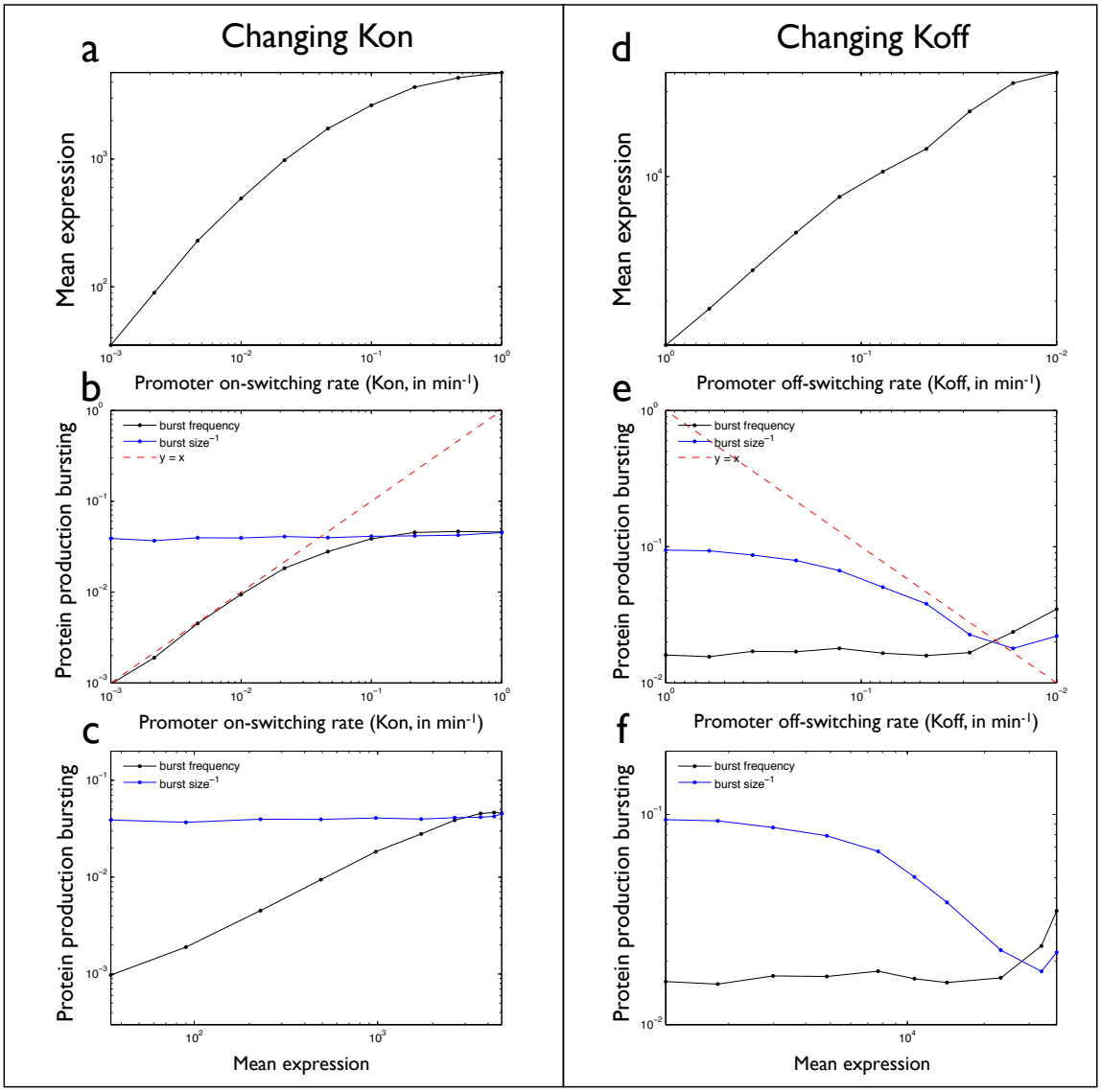

Figure S18: Measured protein production bursting as a function of promoter switching, in a stochastic simulation. We measure mean expression, protein production burst frequency and protein production burst size as a function of changing *Kon* or *Koff*. Increasing *Kon* increases the mean expression level (a) and increases the burst frequency, while burst size remains constant (b). However, there is a regime, at higher *Kon* (b, $Kon>2*10^{-2}$), where measured burst frequency increases non-linearly with increasing *Kon*, while mean expression continues to increase (c). For changing *Koff* we observe a similar phenomenon. Decreasing *Koff* increases the mean expression level (d), but at higher and very low *Koff* (e, $Koff>1*10^{-1}$ and $Koff<2*10^{-2}$) burst size changes non-linearly with *Koff*, while expression continues to increase (f). This analysis shows that changing promoter dynamics can be measured qualitatively through protein production bursting, but might be limited quantitatively in regimes where protein production bursting changes non-linearly with changing promoter dynamics.

Exploring the relationship between mean expression level and noise using an analytical model

We have changed the mean expression and noise using two different biological parameters, namely promoter accessibility (adding polyT) and binding affinity (TF binding site sequence). Using various analyses we show that each biological parameter changes the dynamics of gene expression in a different way. Accessibility changes the frequency of expression bursts and affinity changes the length of expression bursts. Using a stochastic simulation (Gillesple 1977) we show how promoter kinetics (on-off switching) can be modulated to give changes in protein production bursting. Various studies have connected this "bursting" to the noise properties of gene expression. Most notably is the gamma model by Friedman et al. (Friedman et al. 2006), where noise (std^2/mu^2) is inversely related to the burst frequency and the burst size is related to the noise strength (std^2/mu).

To understand how the expression and noise changes that we observe are related to promoter dynamics we use an analytical solution of the kinetic scheme that we used in the stochastic simulations (see below for the derivation), which gives us the mean expression level, noise and noise strength at steady state.

We set the parameters to biologically meaningful values (see above) and assume that changing accessibility changes *Kon* and affinity changes *Koff*, where an increased affinity has a decreased *Koff* and an increased accessibility has an increased *Kon*.

Figure S19 shows the mean versus burst frequency (and noise) plot of either changing *Kon* or *Koff*. Increasing *Kon* increases expression and decreases noise (increases burst frequency) (Fig. S19b,c,e,f blue lines), which is in accordance with our experimental observations (Figure 4 from main text). Interestingly, for changing *Koff* we observe two different qualitative responses, depending on the rate of on switching (*Kon*). When *Kon* is relatively slow decreasing *Koff* increases expression and increases noise (Fig.S19b,c red lines). When *Kon* is relatively fast decreasing *Koff* increases expression and decreases noise (Fig.1S9e,f red lines). In our measurements we observe that increasing the binding affinity increases expression and increases noise. The model therefore predicts that if we would increase *Kon* (TF concentration and/or activity) that noise would decrease with increasing binding affinity. Figure S19d shows expression and burst frequency from gamma fit (inverse noise) for increasing the binding affinity, however now in activating condition (SCD-AA). When starved for amino acids the Gcn4 TF is induced.

Exactly as the model predicts, burst frequency increases (or noise decreases) now with increasing expression.

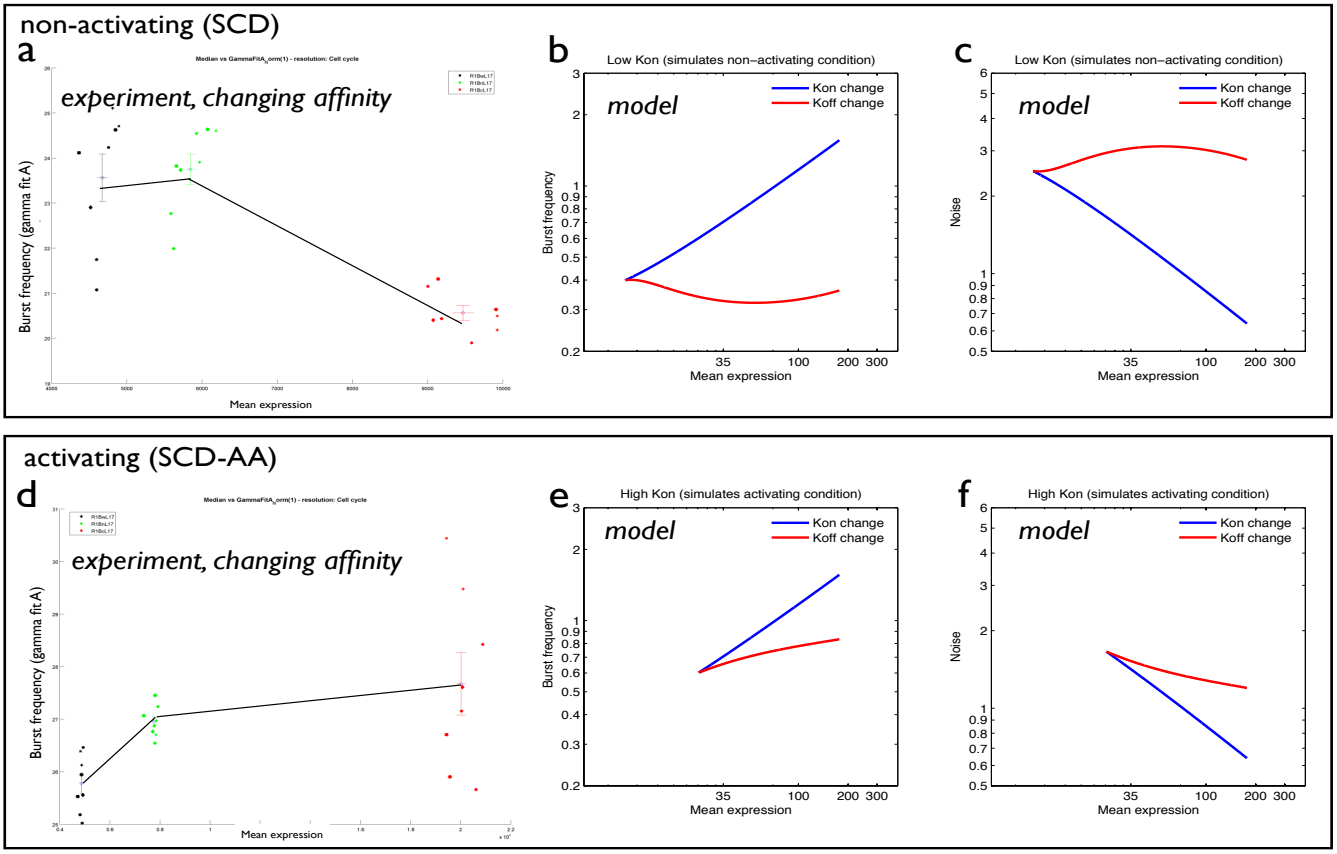

Figure S19: Measured and predicted change in burst frequency and noise. We compare the measured change in expression and burst frequency with the predicted change in expression and burst frequency from the analytical solution of the kinetic scheme shown in figure 1. Burst frequency (a,b,d,e) is quantified as noise^{-1} (c,f) (Friedman et al. 2006). In non-activating condition, where Gcn4 is lowly induced, we measured an increase in noise, or decrease in burst frequency, with increasing expression as a result of increasing the binding affinity of the Gcn4 site (a). The analytical model predicts that such behavior happens when two conditions are met: 1) the promoter is leaky, i.e. the "OFF" state has low, but significant, transcriptional activity, and 2) promoter switching is slow, i.e. *Kon* is relatively low (b,c *Koff* change: red lines). Changing *Kon* in the model always results in an increase in burst frequency and therefore a decrease in noise (b,c,e,f *Kon* change: blue lines). Given these theoretical conditions, the model predicts that the decrease in burst frequency, or increase in noise, disappears when promoter switching is relatively fast, i.e. *Kon* is high (e,f *Koff* change: red lines). In accordance with this prediction, when we repeat the experiment in activating condition (amino acid starved, to induce Gcn4), we find that burst frequency now increases (d), and noise decreases, with increasing expression as a result of increasing the binding affinity.

Frequency spectrum analysis

When the lifetime of mRNAs is short relative to the lifetime of the protein, as is the case for yeast, protein production can be assumed to occur in random uncorrelated events (Cai et al. 2006; Friedman et al. 2006). If protein production is indeed uncorrelated, we don't expect to observe any frequencies, other than cell cycle, in the production over time. **Figure S20** shows the frequency spectrum analysis of the YFP production in time. As expected the only frequency that we observe is around 1/90 per min, which is the frequency of the cell-cycle (measured doubling time). We note however that when mRNA production occurs in bursts, and that re-initiation occurs at some fixed time interval, we would expect to see this in the frequency spectrum. Since re-initiation is probably faster than the interval at which we measure the protein production and due to stochastic translation, we would not be able to pick up this re-initiation frequency.

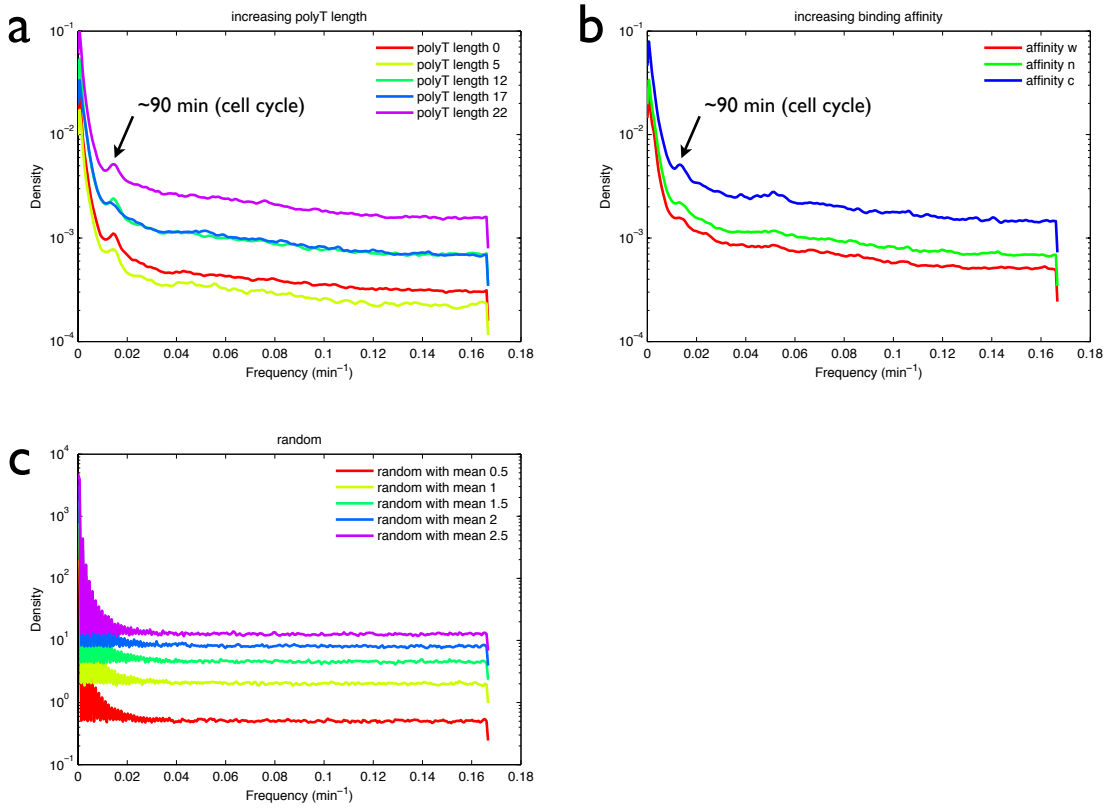

Figure S20: Frequency spectrum analysis of YFP production in time. Shown is a frequency spectrum analysis performed using the Fourier transform of the autocorrelation of the YFP production in time, which was measured in the microscope. The shown spectra quantify the relative contribution of each frequency to the total signal. We find that the only frequency that is significantly present in the data is around 0.01 min^{-1}, which very likely stems from the frequency of the cell cycle (around 90 min, a,b arrow). Changing the polyT length (a) or binding affinity (b) does not change the frequency spectrum. For comparison spectra are shown for random sequences with different means (c).

Analytical solution of the kinetic model

We model stochastic promoter state switching, transcription and translation using the master equation (Sanchez et al. 2011). In this model promoter transcription factor (TF) binding and unbinding events determine the transitions between promoter states. Transcriptional activity changes when the promoter switches to a state with different transcriptional activity. The promoter states can have low (including zero) or high transcription rate, to describe in-active ("off") or active ("on") states respectively. Translation occurs in bursts with the probability of a burst described by a geometric distribution. The master equation (in matrix notation) takes the form:

$$\frac{d}{dt}\vec{p}(n) = \left[\hat{K} - \frac{b}{1+b}\hat{R} - n\delta\hat{I}\right]\vec{p}(n) + \hat{R}\sum_{\beta=1}^{n} h(\beta)\vec{p}(n-\beta) + (n+1)\delta\hat{I}\,\vec{p}(n+1) \quad (1)$$

Where $\vec{p}$ is the vector of probabilities of having *n* proteins in the cell for each promoter state $d/dt\,\vec{p}(n)$ describes the time evolution of these probabilities. $\hat{K}$ is the matrix of promoter state transition rates, where $\hat{K}_{ij}$ is the rate of transitioning from state *j* to state *i* and $\hat{K}_{ii}$ is the sum over all outgoing rates from *i* times *-1*. $\hat{R}$ is the diagonal matrix of transcription rates with $\vec{r}$ on the diagonal ($\hat{R}_{ii} = \vec{r}_i$), where $\vec{r}_i$ is the transcription rate of state *i*. $\hat{I}$ is the identity matrix. *b* is the average burst size (proteins produced per mRNA). δ is the protein degradation rate. *h*(β) describes a geometric distribution and is the probability of producing a burst of size beta.

To derive the mean protein abundance and variance we solve this system at steady state, thus for $d/dt\,\vec{p}(n) = 0$. We get mean protein abundance:

$$\langle n \rangle = \frac{b\,\vec{r}\,\vec{m}_{(0)}}{\delta} \quad (2)$$

Where $\vec{m}_{(0)}$ is the zeroth partial moment of the distribution of mRNA abundance and is the solution to:

$$0 = \hat{K}\,\vec{m}_{(0)} \quad (3)$$

We can get noise (σ^2/μ^2) and noise strength (σ^2/μ) by deriving:

$$\left\langle n^2 \right\rangle = (1+b)\langle n \rangle + \frac{b\,\vec{r}\,\vec{n}_{(1)}}{\delta} \tag{4}$$

Where $\vec{n}_{(1)}$ is the first partial moment of the distribution of protein abundance and is the solution to:

$$0 = (\hat{K} - \delta\,\hat{I})\,\vec{n}_{(1)} + b\,\hat{R}\,\vec{m}_{(0)} \tag{5}$$

Variance (σ^2) is:

$$Var(n) = \left\langle n^2 \right\rangle - \langle n \rangle^2 \tag{6}$$

Therefore noise (σ^2/μ^2) becomes:

$$\left\langle \eta^2 \right\rangle = \frac{(1+b)\langle n \rangle - \langle n \rangle^2 + \frac{b\,\vec{r}\,\vec{n}_{(1)}}{\delta}}{\langle n \rangle^2} \tag{7}$$

And noise strength (σ^2/μ):

$$\langle F \rangle = \frac{(1+b)\langle n \rangle - \langle n \rangle^2 + \frac{b\,\vec{r}\,\vec{n}_{(1)}}{\delta}}{\langle n \rangle} \tag{8}$$

We describe gene expression using a two state kinetic scheme that represents switching between an active (ON) and inactive (OFF) promoter configuration (**Figure S16**). The ON state is transcriptionally active (with rate *r1*) and we allow the OFF state to have some (leaky) transcriptional activity (with rate *r2*). $\hat{K}$ and $\vec{r}$ thus become:

$$\hat{K} = \begin{pmatrix} -Koff & Kon \\ Koff & -Kon \end{pmatrix} \tag{9}$$

$$\vec{r} = \begin{pmatrix} r1 & r2 \end{pmatrix} \tag{10}$$

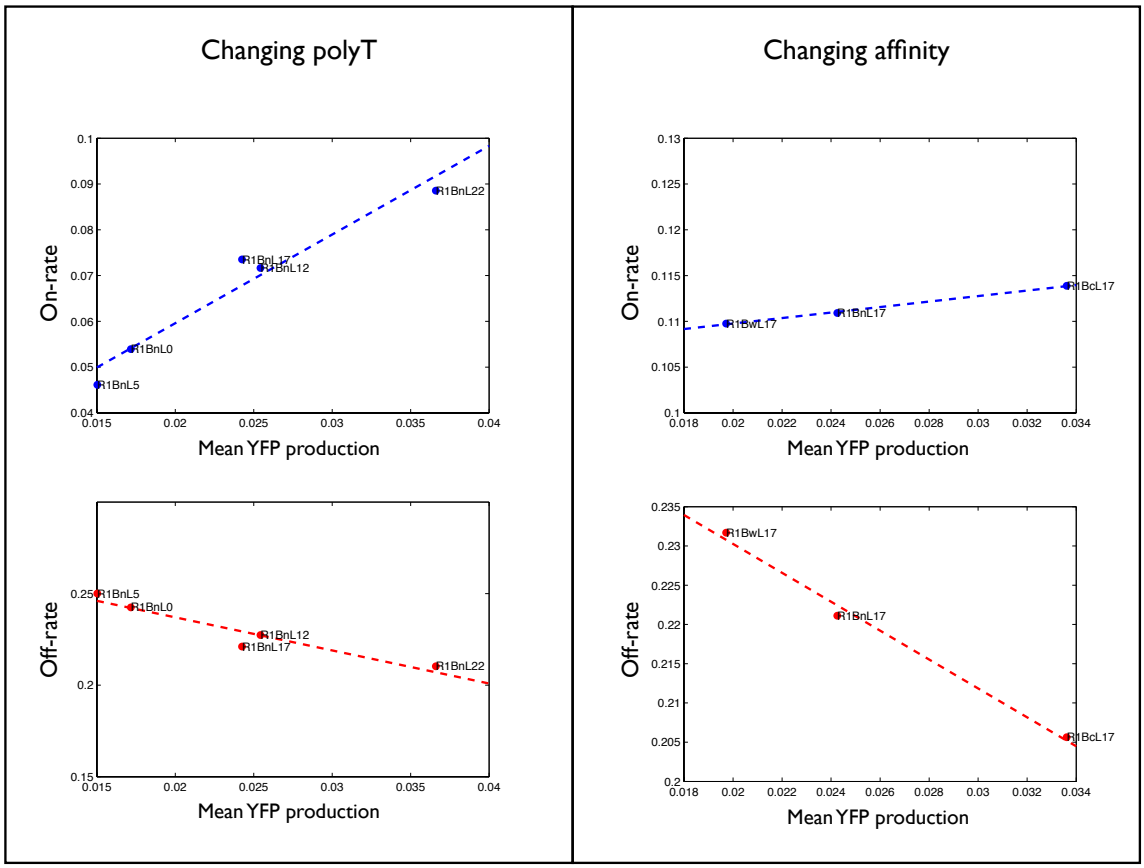

Figure S21: Analysis of protein production bursting using the change in production rate. We define a production burst (ON state) as a period in which the change in YFP production is positive, i.e. the second derivative of the YFP fluorescence in time. We then quantify the on-rate as the number of OFF to ON transitions divided by the total time in the OFF state. Similarly we quantify the off-rate as the number of ON to OFF transitions divided by the total time in the ON state. Shown is the measured off and on-rate (from microscope data) versus the mean production rate, for either changing the polyT length or the binding site affinity. We find that when changing the polyT length the on-rate changes more than the off-rate. When changing the binding affinity we find that the off-rate changes more than the on-rate. Measured rates are in min^{-1}.

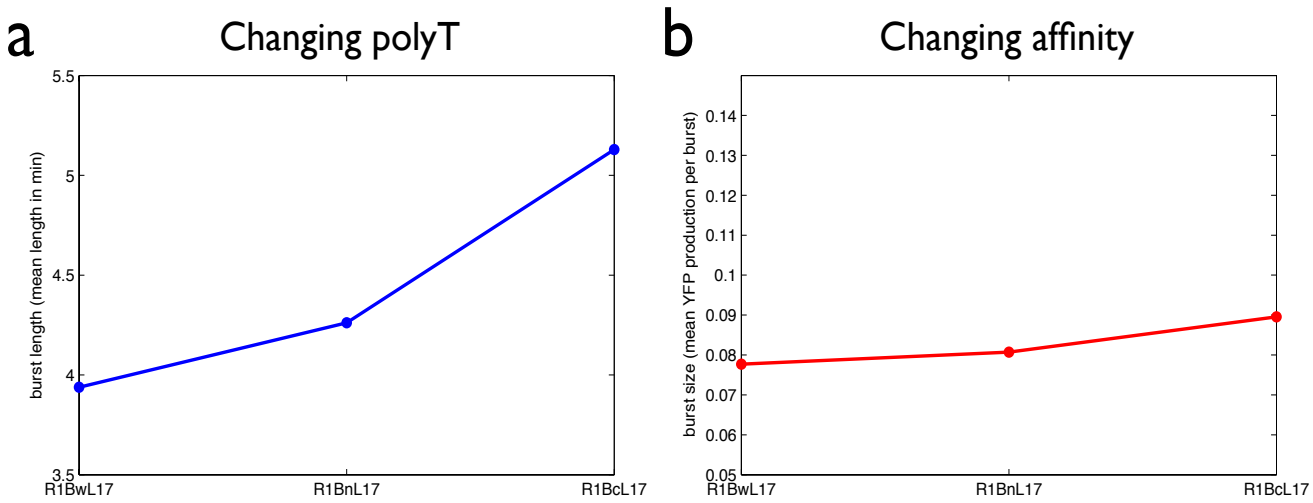

Figure S22: Burst size and burst length as a function of changing the binding site affinity. We define a "burst" as a period of relative high (YFP) protein production. Shown is the average burst length and average burst size, measured in microscope, for strains with differing binding site affinity. We quantified the length, in minutes, of the periods of activity (a) and the size in terms of average YFP production (b).

Supplemental References

Cai L, Friedman N, Xie XS. 2006. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**: 358–62. http://www.ncbi.nlm.nih.gov/pubmed/16541077 (Accessed July 16, 2011).

Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin D a, Chang JH, Lindquist R a, Moffat J, et al. 2006. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology* **7**: R100. http://www.ncbi.nlm.nih.gov/pubmed/17076895.

Cookson S, Ostroff N, Pang WL, Volfson D, Hasty J. 2005. Monitoring dynamics of single-cell gene expression over multiple cell cycles. *Molecular systems biology* **1**: 2005.0024. http://www.ncbi.nlm.nih.gov/pubmed/16729059.

Friedman N, Cai L, Xie X. 2006. Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression. *Physical Review Letters* **97**: 1–4. http://link.aps.org/doi/10.1103/PhysRevLett.97.168302 (Accessed August 9, 2011).

Gillesple DT. 1977. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry* **81**: 2340–2361.

Ginigea E, Ptashne M. 1988. DNA binding of the yeast transcriptional activator GAL4. *Biochemistry* **85**: 382–386.

von der Haar T. 2008. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. *BMC systems biology* **2**: 87. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2590609&tool=pmcentrez&rendertype=abstract (Accessed November 16, 2012).

Liang SD, Marmorstein R, Harrison SC, Ptashne M. 1996. DNA sequence preferences of GAL4 and PPR1: how a subset of Zn2 Cys6 binuclear cluster proteins recognizes DNA. *Molecular and cellular biology* **16**: 3773–80. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=231373&tool=pmcentrez&rendertype=abstract.

Octavio LM, Gedeon K, Maheshri N. 2009. Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression. ed. A. Akhtar. *PLoS genetics* **5**: e1000673. http://dx.plos.org/10.1371/journal.pgen.1000673 (Accessed June 5, 2011).

Pelechano V, Chavez S, Perez-Ortin JE. 2010. A Complete Set of Nascent Transcription Rates for Yeast Genes. *PloS one* **5**: 1–10.

Raveh-Sadka T, Levo M, Shabi U, Shany B, Keren L, Lotan-Pompan M, Zeevi D, Sharon E, Weinberger A, Segal E. 2012. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nature genetics* **44**: 743–50. http://www.ncbi.nlm.nih.gov/pubmed/22634752 (Accessed November 14, 2012).

Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J. 2011. Effect of promoter architecture on the cell-to-cell variability in gene expression. *PLoS computational biology* **7**: e1001100. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3048382&tool=pmcentrez&rendertype=abstract (Accessed July 22, 2011).

Tan RZ, van Oudenaarden A. 2010. Transcript counting in single cells reveals dynamics of rDNA transcription. *Molecular systems biology* **6**: 358. http://www.ncbi.nlm.nih.gov/pubmed/20393578 (Accessed November 22, 2012).

Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO. 2002. Precision and functional specificity in mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 5860–5. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=122867&tool=pmcentrez&rendertype=abstract.

Zenklusen D, Larson DR, Singer RH. 2011. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature structural & molecular biology* **15**: 1263–1271.